EXHIBIT 1

PLAINTIFFS' DISPUTED RFPS TO DEFENDANT GOOGLE LLC

1 **PROPOUNDING PARTIES:**   **Consolidated Plaintiffs**

2 **RESPONDING PARTY:**   **Defendant Google LLC**

3 **SET NUMBERS:**   **One and Two**

4   Pursuant to Rules 26 and 34 of the Federal Rules of Civil Procedure, Plaintiffs in the above-titled

5 action hereby request that Defendant Google LLC produce to Consolidated Plaintiffs all Documents,

6 electronically stored information, and things that are in Your possession, custody, or control responsive to

7 the following Requests for Production.

8   **DEFINITIONS**

9   1.   Each of these definitions and instructions is incorporated into each of the requests to which

10 it pertains. Notwithstanding any definition below, each word, term, or phrase used in these requests for

11 production is intended to have the broadest meaning permitted under the Federal Rules of Civil Procedure.

12   2.   The terms "relating to," "related to," "relate to," "refer to," or "concerning" or any variations

13 thereof, mean referring to, pertaining to, reflecting, containing, evidencing, supporting, describing,

14 establishing, showing, disclosing, explaining, mentioning, analyzing, constituting, comprising, setting

15 forth, summarizing, characterizing, and/or contradicting, either directly or indirectly, in full or in part.

16   3.   "All," "or," and "and" should be understood to include and encompass "any"; "or" should

17 be understood to include and encompass "and"; and "and" should be understood to include and encompass

18 "or."

19   4.   "Action" means this consolidated action, captioned *In Re Google Generative AI Litigation*,

20 Master Case No. 5:23-cv-03440-EKL (N.D. Cal.)

21   5.   "Agent" means any employee, officer, manager, director, attorney, representative,

22 independent contractor, subcontractor, day laborer, temporary worker, or any other person—regardless of

23 formality of relationship—acting at Your direction or on Your behalf.

24   6.   "AI-Powered Products" means Google products that integrate and/or are powered by one or

25 more Generative AI Models.

26   7.   "Alphabet" means Defendant Alphabet Inc., together with any of its corporations,

27 businesses, subsidiaries, divisions, subdivisions, predecessors, successors, parents, and affiliates, and their

28 respective officers, directors, employees, partners, representatives, Agents, attorneys, accountants, or other

Master File Case No.: 5:23-cv-03440-EKL   1

PLAINTIFFS' DISPUTED REQUESTS FOR PRODUCTION TO DEFENDANT GOOGLE LLC

1   persons occupying similar positions or performing similar functions.

2   8.   "And" and "or" shall be construed conjunctively or disjunctively as necessary to make the

3   Request inclusive rather than exclusive.

4   9.   "Artificial Intelligence" or "AI" engineered system that generates outputs such as content,

5   forecasts, recommendations, or decisions for a given set of human-defined objectives.

6   10.   Unless otherwise specified in a Request, "Bard" refers to the large language model ("LLM")

7   (i.e., AI software designed to emit convincingly naturalistic text outputs in response to user prompts)

8   created, maintained, distributed, and/or sold by Google, including all versions, iterations, and/or model

9   family members of Bard

10   11.   "C4" means the dataset created by Google and referenced in the February 10, 2022 research

11   paper entitled LAMDA: Language Models for Dialog Applications.

12   12.   "Common Crawl" means the nonprofit organization that crawls the web and makes its

13   archives and datasets publicly available.

14   13.   "Communication" means every exchange of information of any nature, whether oral,

15   written, or recorded, from one person or entity to another, and any evidence of such exchange, Including,

16   but not limited to, any correspondence, memorandum, electronic mail messages, social media postings,

17   messages sent or received through social media, collaboration platforms or other online service such as

18   Slack, Workplace, Teams, message board postings, text messages, letters, notes or logs of meetings,

19   diaries, daily calendars, videos, audio recordings, or other records of exchanges between people or entities.

20   14.   "Consolidated Plaintiffs" means Plaintiffs Steve Almond, Sarah Andersen, Burl Barer,

21   Jessica Fink, Kirsten Hubbard, Hope Larson, Mike Lemos, Jill Leovy, Connie McLennan, and Jingna

22   Zhang.

23   15.   "Google" means Defendant Google LLC, together with any of its corporations, businesses,

24   subsidiaries, divisions, subdivisions, predecessors, successors, parents, and affiliates, and their respective

25   officers, directors, employees, partners, representatives, Agents, attorneys, accountants, or other persons

26   occupying similar positions or performing similar functions.

27   16.   "Document" means any document or electronically stored information as described in

28   Federal Rule of Civil Procedure 34(a) Including, but not limited to, any handwriting, typewriting, printing,

PLAINTIFFS' DISPUTED REQUESTS FOR PRODUCTION TO DEFENDANT GOOGLE LLC

photostating, photographing, photocopying, transmitting by electronic mail or facsimile, and every other means of recording upon any tangible thing, any form of Communication or representation, Including letters, words, pictures, sounds, or symbols, or combinations thereof, and any record thereby created, regardless of the manner in which the record has been stored.

This Includes:

- The originals, drafts and All non-identical copies thereof, whether different from the original by reason of any notation made on such copies or otherwise;

- Booklets, brochures, pamphlets, circulars, notices, periodicals, papers, contracts, agreements, photographs, minutes, memoranda, messages, appraisals, analyses, reports, financial calculations and representations, invoices, accounting and diary entries, inventory sheets, diaries, appointment books or calendars, teletypes, telefaxes, thermafaxes, ledgers, trial balances, correspondence, telegrams, press releases, advertisements, notes, working papers, drawings, schedules, tabulations, projections, information or programs stored in a computer (whether or not ever printed out or displayed), and All drafts, alterations, modifications, changes or amendments of any of the foregoing;

- Graphic or aural representations of any kind, Including, without limitation, photographs, microfiche, microfilm, videotapes, recordings, drawing, charts and motion pictures;

- All letters, words, pictures, sounds, or symbols, or combinations thereof stored in or on any electronic, mechanical, magnetic, or optical device Including, but not limited to: (i) computer data storage devices (servers, laptops hard-drives, flash drives, discs, magnetic cards, and the like), (ii) the internet or "the Cloud" (such as e-mail, web posts, social media posts, internet pages, or Google Suite documents (e.g., Google Docs or Google Sheets) etc.), and (iii) information stored on cell phones.

17.   Unless otherwise specified in a Request, "Gemini" refers to the multimodal large language model (see *infra*) created, maintained, distributed, and/or sold by Google, including all versions, iterations, and/or model family members of Gemini such as Gemini 1.0, Gemini 1.0 Pro, Gemini 1.0 Ultra, Gemini 1.0 Nano, Gemini 1.5 Pro, Gemini 1.5 Flash, Gemini 1.5 Flash-8B, Gemini 2.0, Gemini 2.0 Flash, and the Gemini App.

18.   "Identify," with respect to a Document means, to give, to the extent known, the (i) type of Document (*i.e.*, correspondence, memorandum, business record, etc.); (ii) general subject matter; (iii) date of the Document; (iv) author(s) or preparer(s), addressee(s), and recipient(s) of each such Document; (v) the present location of each such Document or copies thereof; and (vi) if applicable, the Bates number of the Document. "Identify," with respect to an entity, governmental body, department, or agency (not a natural Person), means to state the full name, address, and telephone number for its principal place of business. "Identify," with respect to a natural Person means, to give, to the extent known, the (i) Person's full name and title; (ii) current or last known address; (iii) current or last known business and telephone number(s); and (iv) current or last known place of employment. Once a Person has been identified in accordance with this paragraph, only the name of that Person need be listed in response to subsequent discovery requesting the identification of such Person.

19.   "Including, "including, "include," or "includes" is used to provide examples of certain types of information and should not be construed as limiting a request or definition in any way. The term "Including, " "including," "include," or "includes" shall be construed as if followed by the phrase "but not limited to."

20.   "Information" or "information" means written, recorded, graphic, or other electronic data, including metadata, of any nature whatsoever, regardless of how recorded, and whether original or copy.

21.   Unless otherwise specified in a Request, "Imagen" refers to the text-to-image diffusion model (i.e., an AI software product designed to generate an image in response to a text prompt) created, maintained, distributed, and/or sold by Google, including all versions, iterations, and/or model family members of Imagen.

22.   "Image-Text Pair" means a dataset element comprising both an image and its corresponding textual description, used to train AI models on understanding and generating content that combines visual and textual information.

23.   "Infiniset" means the dataset created by Google and referenced in the February 10, 2022 research paper entitled *LAMDA: Language Models for Dialog Applications.*

24.   "LAION" means Large-Scale Artificial Intelligence Open Network (an organization based in Hamburg, Germany), its predecessor and successor entities, its officers, directors, engineers, researchers, employees, agents, attorneys, representatives, and other persons acting or authorized to act on its behalf.

25.   "LAION Datasets" includes all datasets publicly released or privately developed by LAION, whether alone or in collaboration with others.

26.   "LLM" or "large language model" means an AI system trained on large amounts of text data to understand and generate human-like language.

27.   "MLLM" or "multimodal large language model" means an LLM that accepts input and generates output in multiple modalities and is trained using text, image, audio, and video data.

28.   "Other Generative AI Models" means all generative AI models created, maintained, distributed, and/or sold by Google LLC other than Imagen, Bard, and Gemini, including but not limited to, all versions, iterations, and/or model family members of PaLM, GLaM, LaMDA, Codey, Chirp, Veo, MedLM, LearnLM, SecLM, Gemma, Code Gemma, Recurrent Gemma, PaliGemma.

29.   "Person" means any natural persons, proprietorships, legal or governmental entities, corporations, partnerships, trusts, joint ventures, groups, associations, or organizations. Any reference to a "person" shall mean that "person" and all affiliates, divisions, controlled companies, subsidiaries, or otherwise related entities and all of his, her, or its current and former officers, directors, employees, agents, representatives, attorneys, and accountants.

30.   "Plaintiffs" refers to the plaintiffs in the above-entitled action.

31.   "Such as" or "such as" is used to provide examples of certain types of information and should not be construed as limiting a request or definition in any way. The term "Such as" or "such as" shall be construed as if followed by the phrase "but not limited to."

32.   "Training Data" refers to any data used to develop, pre-train, train, test, fine-tune, or modify any AI, machine learning, or deep learning model, Including any textual, visual, audio, or multimedia data, or metadata related thereto. "Training Data" Includes All datasets, structured and unstructured data, images, audio-visual works, textual descriptions, and metadata whether obtained through scraping, licensed sources, third-party datasets (Including LAION-400M), internal datasets, or

PLAINTIFFS' DISPUTED REQUESTS FOR PRODUCTION TO DEFENDANT GOOGLE LLC

1   any other source. To the extent there are multiple variations, transformations, and/or iterations of such

2   data, "Training Data" refers to each separate variation, transformation, and iteration of such data.

3   33.   "You" or "Your" refers to and Includes the specific Defendant producing Documents in

4   response to these Requests, its employees, agents, attorneys, accountants, representatives, predecessors or

5   successors-in-interest, any corporation or partnership under its direction, and any other person or entity

6   acting on its behalf or under its control.

7   **INSTRUCTIONS**

8   1.   If, in responding to these requests, You encounter ambiguities when construing a question,

9   instruction, or definition, Your answer shall set forth the matter deemed ambiguous and the construction

10   used in answering.

11   2.   Unless otherwise specified, Documents (Including electronically stored information) shall

12   be produced consistent with and in the manner set forth in the applicable ESI Protocol or, to the extent such

13   ESI Protocol has not yet been entered by the Court, the portions of the ESI Protocol upon which the parties

14   agree.

15   3.   All Documents retained in a machine-readable format should be produced in that machine-

16   readable format.

17   4.   These Requests for Productions should be deemed continuing such that if Your directors,

18   officers, employees, agents, representatives, or any person acting on Your behalf, subsequently discover or

19   obtain possession, custody, or control of any Document or ESI previously requested or required to be

20   produced, and supplemental productions should be provided as additional Documents become available.

21   5.   If You claim You are unable to produce a Document, you must state whether that inability

22   is because the Document never existed; has been destroyed, lost, misplaced or stolen; or has never been or

23   is no longer in your possession, custody or control. Such a statement must further set forth the name and

24   address of any person or entity that you know or believe to have possession, custody or control of that item

25   or category of item. If any Document responsive to a request has been destroyed, produce All Documents

26   describing or referencing (1) the contents of the lost or destroyed Document; (2) All locations in which any

27   copy of the lost or destroyed Document had been maintained; (3) the date of any such loss or destruction

28   to the extent known; (4) the name of each person who ordered, authorized and carried out the destruction

of any lost or destroyed Document; (5) All Document retention and destruction policies in effect at the time any requested Document was destroyed; and (6) All efforts made to locate any responsive Document alleged to have been lost or destroyed.

6.   If You object to any item or category of item, Your response shall (a) identify with particularity each Document or thing to which the objection is made and (b) set forth clearly the extent of, and specific ground for, the objection; and YOU should respond to the Request to the extent it is not objectionable.

7.   If You object that a Document is covered by the attorney-client or other privilege, or is work-product, You shall provide a Privilege Log containing: (1) the name of the Document; (2) the name and address of the person(s) who prepared it; (3) the person(s) to whom it was directed or circulated; (4) the date on which the Document was prepared or transmitted; (5) the name and address of the person(s) now in possession of the Document; (6) the description of the subject matter of the Document; and (7) the specific nature of the privilege claimed, Including the reasons and each and every fact supporting the withholding, and legal basis sufficient to determine whether the claim of privilege is valid with respect to the Document (without revealing privileged information).

8.   In responding to these Requests for Production, You are to Include Documents (1) obtained from witnesses who gave information to any governmental agency or investigatory body, Including Congress; (2) that constitute, refer or relate to summaries of testimony or other statements in connection with any governmental agency or investigatory body proceeding or investigation, Including before Congress; or (3) obtained on Your behalf in preparation for testimony or interviews before any governmental agency or investigatory body, Including Congress.

9.   Unless otherwise specified, the Relevant Time Period is from January 1, 2017 until the Present.

REQUESTS FOR PRODUCTION

REQUEST FOR PRODUCTION NO. 4:

All Documents and Communications related to Google's internal technical & engineering documentation of Imagen, Bard, and Gemini, including system design, training, deployment, and operations.

REQUEST FOR PRODUCTION NO. 5:

All Documents and Communications related to Google's internal tools used to store, analyze, catalog, and retrieve Training Data.

REQUEST FOR PRODUCTION NO. 6:

All Documents and Communications related to Google's internal data dictionaries, database field descriptions, database field mappings, and developer manuals, including all drafts, concerning Imagen, Bard, and Gemini.

REQUEST FOR PRODUCTION NO. 13:

All Documents and Communications concerning the processes, methodologies, or procedures for collecting, organizing, curating, preprocessing, or annotating the Training Data used or considered for use in Imagen, Bard, and Gemini. This Includes records regarding decisions on data selection, filtering, usage of classifiers (for detection of copyrighted material), exclusion, augmentation, labeling; change-logs and programming notes (in a machine-readable format) reflecting the implementation of such decisions; and any involvement of third-party services, tools, or contractors in these processes.

REQUEST FOR PRODUCTION NO. 28:

All Documents and Communications, Including training logs, version histories, changelogs, or other records, showing updates, modifications, or changes made to Imagen, Bard, and Gemini such as but not limited to git repository commit messages, pull requests, code base, and magnet links. This Includes records of any changes implemented to address, mitigate, or avoid potential copyright infringement, as well as documentation of decisions made based on those concerns.

REQUEST FOR PRODUCTION NO. 29:

All Documents and Communications sufficient to show the development, evaluation, testing, or integration of content management or filtering systems (e.g., YouTube Content ID) for Imagen, Bard, and Gemini.

REQUEST FOR PRODUCTION NO. 35:

All Documents and Communications related to Google's internal technical & engineering documentation of the Other Generative AI Models, including system design, training, deployment,

and operations.

REQUEST FOR PRODUCTION NO. 36:

All Documents and Communications related to Google's internal data dictionaries,
database field descriptions, database field mappings, and developer manuals, including all drafts,
concerning the Other Generative AI Models.

REQUEST FOR PRODUCTION NO. 42:

All Documents and Communications concerning the processes, methodologies, or procedures
for collecting, organizing, curating, preprocessing, or annotating the Training Data used or considered
for use in the Other Generative AI Models. This Includes records regarding decisions on data
selection, filtering, exclusion, augmentation, labeling; change-logs and programming notes (in a
machine-readable format) reflecting the implementation of such decisions; and any involvement of
third-party services, tools, or contractors in these processes.

REQUEST FOR PRODUCTION NO. 53:

All Documents and Communications, Including training logs, version histories, changelogs,
or other records, showing updates, modifications, or changes made to the Other Generative AI Models
such as but not limited to git repository commit messages, pull requests, code base, and magnet links.
This Includes records of any changes implemented to address, mitigate, or avoid potential copyright
infringement, as well as documentation of decisions made based on those concerns.

REQUEST FOR PRODUCTION NO. 54:

All Documents and Communications sufficient to show the development, testing, evaluation,
or integration of content management or filtering systems (e.g., YouTube Content ID) for the Other
Generative AI Models.

REQUEST FOR PRODUCTION NO. 60:

All policies and procedures, including all iterations of such policies and procedures related to
identifying, obtaining, downloading, assessing, vetting and selecting sources of training data for Your
Generative AI Models and/or any of Your AI-Powered Product(s), including any means identified in
response to Interrogatory No. 5.

Master File Case No.: 5:23-cv-03440-EKL   9

REQUEST FOR PRODUCTION NO. 61:

All policies and procedures, including all iterations of such policies and related to the use and treatment of copyrighted or potentially copyrighted material contained within the training data for Your Generative AI Models and/or any of Your AI-Powered Product(s).

REQUEST FOR PRODUCTION NO. 62:

All policies, procedures, systems, and other measures that You implemented between January 1, 2017 and the present to ensure compliance with United States copyright laws in connection with selecting, obtaining, copying, processing, and using training data for Your Generative AI Models, including any measures for: identifying copyrighted works before copying them; obtaining necessary authorizations from rights holders; tracking which works were copied; preventing unauthorized copying; removing works upon receiving takedown notices; preventing verbatim reproduction of protected content; monitoring ongoing compliance; and evaluating fair use.

REQUEST FOR PRODUCTION NO. 63:

All policies and procedures that You follow to assess risk, safety, and alignment, including the risk that a version or iteration will output protected expression from Training Data, before You release a version or iteration of Your Generative AI Models and/or any of Your AI-Powered Product(s) to the public, including policies and procedures describing the risks to be assessed.